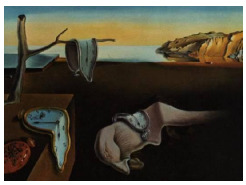
View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Journal of Statistical Software



Journal of Statistical Software
April 2009, Volume 30, Issue 7. http://www.jstatsoft.org/

MetaEasy: A Meta-Analysis Add-In for Microsoft Excel

Evangelos Kontopantelis
National Primary Care Research and Development Centre

David Reeves
National Primary Care Research and Development Centre

Abstract

Meta-analysis is a statistical methodology that combines or integrates the results of several independent clinical trials considered by the analyst to be 'combinable' (Huque 1988). However, completeness and user-friendliness are uncommon both in specialised meta-analysis software packages and in mainstream statistical packages that have to rely on user-written commands. We implemented the meta-analysis methodology in an Microsoft **Excel** add-in which is freely available and incorporates more meta-analysis models (including the iterative maximum likelihood and profile likelihood) than are usually available, while paying particular attention to the user-friendliness of the package.

Keywords: meta-analysis, forest plot, **Excel**, VBA, maximum likelihood, profile likelihood.

1. Introduction

Meta-analysis can be defined as the statistical analysis of a large collection of analysis results from individual studies - usually Randomised Controlled Trials (RCTs) - for the purpose of integrating the findings (Glass 1976). Although the debate regarding the quality and application caveats of the method is ongoing (Egger and Smith 1997; Bailar 1997), a Medline (`http://medline.cos.com/`) search by the authors reveals that the number of meta-analysis studies published in peer-reviewed journals seems to be growing exponentially (Figure 1). Published meta-analysis studies (search criterion: Publication Type=meta-analysis) have risen from 274 in 1990 to 2138 in 2005, while published work that is either a meta-analysis or deals with meta-analysis issues (search criterion: Keyword=meta-analysis) has increased from 329 to 3350, in the same period.

A major issue in meta-analysis is the almost inevitable clinical or methodological heterogeneity among the combined studies (Eysenck 1994). If the study results differ greatly (large

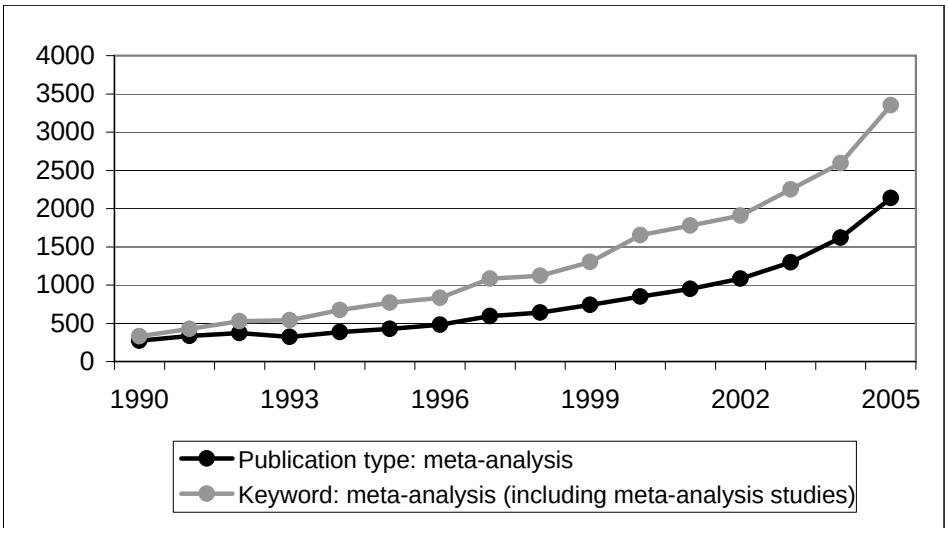


Figure 1: Number of meta-analysis publications (Medline search).

heterogeneity) then it may not be appropriate to combine them, since an estimate of overall effect in such a case is of questionable value (Egger, Smith, and Phillips 1997; Thompson 1994). Nevertheless, the statistical error that stems from this between-study diversity can be quantified and modelled (Higgins and Thompson 2002; DerSimonian and Laird 1986), and a meta-analysis on carefully selected Randomised Controlled Trials, in terms of compatibility, is the next best thing to a new large and expensive, prohibitively so in many cases, RCT.

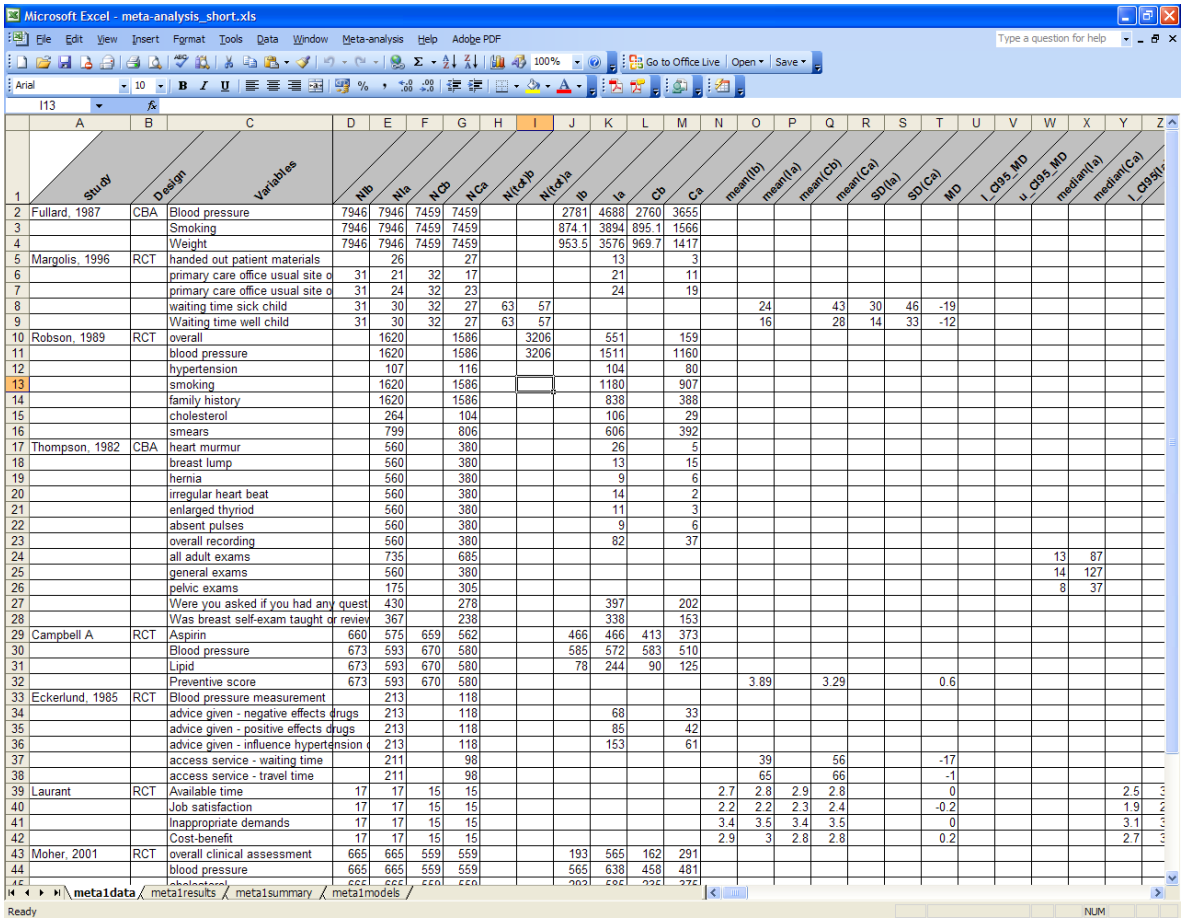
Although SAS (SAS Institute Inc 2003), Stata (StataCorp. 2007), and SPSS (SPSS Inc 2006) do not include embedded meta-analysis commands, some user-written commands and/or macros exist that deal with this shortcoming - some more successfully than others (Egger, Sterne, and Smith 1998). Alternatively, a researcher may use more user-friendly specialist meta-analysis software packages, like **MetaWin** (Rosenberg, Adams, and Gurevitch 2007), the free but potent **RevMan** (The Cochrane Collaboration 2008) or the new, promising and also free **MIX** (Bax, Yu, Ikeda, Tsuruta, and Moons 2006). However, in most of the above, the outcomes of the included studies that will be 'fed' into the models must all be expressed in identical format (eg group means and standard deviations; odds ratios and sample sizes) for all the studies. Therefore, the researcher must go through a preliminary process of transforming study outcomes that may have been disseminated using a variety of statistical parameters (eg means and *SD*s; t-values; p-values) to the common format, a tedious task that requires at least some statistical adeptness.

Since the meta-analyst will often use a spreadsheet to summarise the reported study outcomes and other details, in order to better organise the analysis, we developed an add-in for Microsoft **Excel** that automates many meta-analysis processes and provides support for the task. The first purpose of our software is to calculate an effect size and its standard error, from the specific combination of input parameters supplied by the user for each outcome, using one of the methods described by the Cochrane Collaboration (Higgins and Green 2006). The methods that can be used with each set of input parameters are automatically identified and the one that produces the most precise estimate of the effect is selected. Following that,

a forest plot is created (Lewis 2001) summarising all the outcomes organised by study. Finally, various meta-analysis models, the more advanced of which (maximum likelihood, profile likelihood and permutations method) are not available in any other meta-analysis software package even though they have been proved to be more robust for normally distributed effects (Brockwell and Gordon 2001; Follmann and Proschan 1999), are used to calculate an overall mean effect and its variance. Results of the models are displayed in a second forest plot, while a variety of heterogeneity measures are provided to help the user decide on the appropriate model for the analysis at hand.

2. Data entry

After the add-in has been installed a meta-analysis menu will be available on the Microsoft **Excel** menu-bar. Once the worksheets have been formatted using the `Format Sheets` command, data can be entered on the first sheet (Figure 2) - a maximum of 10 meta-analyses can be accommodated in a workbook but we will refer to a single one for simplicity. The fields on the data-entry worksheet are described in Table 1.



	Study	Design	Variables	N!b	N!a	N!Cb	N!Ca	N(t|d)b	N(t|d)a	!b	!a	Cb	Ca	mean(!b)	mean(!a)	mean(Cb)	mean(Ca)	SD(!a)	SD(Ca)	MD	l_CI95_MD	u_CI95_MD	median(!a)	median(Ca)	l_CI95(C
2	Fullard, 1987	CBA	Blood pressure	7946	7946	7459	7459			2781	4688	2760	3655												
3			Smoking	7946	7946	7459	7459			874.1	3894	895.1	1566												
4			Weight	7946	7946	7459	7459			953.5	3576	969.7	1417												
5	Margolis, 1996	RCT	handed out patient materials		26		27				13		3												
6			primary care office usual site o	31	21	32	17				21		11												
7			primary care office usual site o	31	24	32	23				24		19												
8			waiting time sick child	31	30	32	27	63	57						24		43	30	46	-19					
9			Waiting time well child	31	30	32	27	63	57						16		28	14	33	-12					
10	Robson, 1989	RCT	overall		1620		1586		3206		551		159												
11			blood pressure		1620		1586		3206		1511		1160												
12			hypertension		107		116				104		80												
13			smoking		1620		1586				1180		907												
14			family history		1620		1586				838		388												
15			cholesterol		264		104				106		29												
16			smears		799		806				606		392												
17	Thompson, 1982	CBA	heart murmur		560		380				26		5												
18			breast lump		560		380				13		15												
19			hernia		560		380				9		6												
20			irregular heart beat		560		380				14		2												
21			enlarged thyriod		560		380				11		3												
22			absent pulses		560		380				9		6												
23			overall recording		560		380				82		37												
24			all adult exams		735		685																	13	87	
25			general exams		560		380																	14	127	
26			pelvic exams		175		305																	8	37	
27			Were you asked if you had any quest	430		278					397		202												
28			Was breast self-exam taught or reviev	367		238					338		153												
29	Campbell A	RCT	Aspirin	660	575	659	562			466	466	413	373												
30			Blood pressure	673	593	670	580			585	572	583	510												
31			Lipid	673	593	670	580			78	244	90	125												
32			Preventive score	673	593	670	580								3.89		3.29			0.6					
33	Eckerlund, 1985	RCT	Blood pressure measurement		213		118																		
34			advice given - negative effects drugs		213		118				68		33												
35			advice given - positive effects drugs		213		118				85		42												
36			advice given - influence hypertension		213		118				153		61												
37			access service - waiting time		211		98								39		56			-17					
38			access service - travel time		211		98								65		66			-1					
39	Laurant	RCT	Available time	17	17	15	15							2.7	2.8	2.9	2.8			0					2.5
40			Job satisfaction	17	17	15	15							2.2	2.2	2.3	2.4			-0.2					1.9
41			Inappropriate demands	17	17	15	15							3.4	3.5	3.4	3.5			0					3.1
42			Cost-benefit	17	17	15	15							2.9	3	2.8	2.8			0.2					2.7
43	Moher, 2001	RCT	overall clinical assessment	665	665	559	559			193	565	162	291												
44			blood pressure	665	665	559	559			565	638	458	481												

Figure 2: Data sheet.

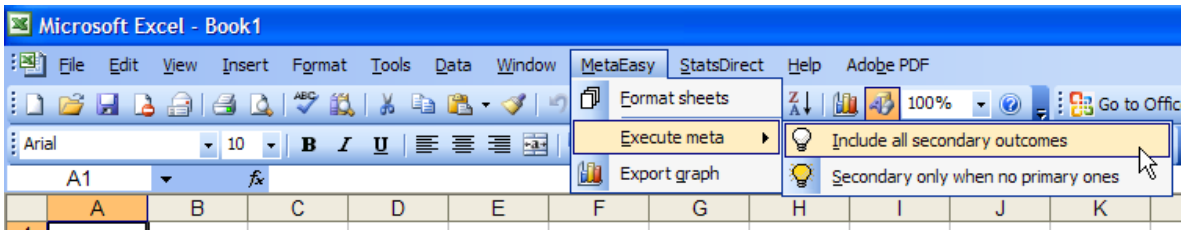


Figure 3: Meta-analysis add-in menu.

Once data has been inputted all processes are executed by selecting one of the two available meta-analysis options that will be explained in more detail later: `Include all secondary outcomes` or `Secondary only when no primary ones` (Figure 3).

Name	Label	Type	Required	Information
Study	The name of the study	String	Yes	Only needs to be inputted once for each study (in the first outcome row). If the colour background of a cell on that column is set to red, the row/outcome is not included in the analysis
Design	The type of the design (RCT, Observational study,etc)	String	No	Only needs to be inputted once for each study (in the first outcome/variable row)
Variables	Outcome variable names	String	Yes	Truncated to 28 characters in graphs. It is advised to set the names of primary outcomes in bold and those of secondary outcomes in italics. Outcomes with names in italics are plotted separately in the outcomes scatter plot
NIb	Intervention group size, before treatment	Integer	–	Not used in any of the methods, column provided for information purposes and/or future use
NIa	Intervention group size, after treatment	Integer	Yes	
NCb	Control group size, before treatment	Integer	–	Not used in any of the methods, column provided for information purposes and/or future use
NCa	Control group size, after treatment	Integer	Yes	
$N(tot)b$	$NIb + NCb$	Integer	–	Not used in any of the methods, column provided for information purposes

table continued to next page

Name	Label	Type	Required	Information
$N(tot)a$	$NIa + NCa$	Integer	–	Not used in any of the methods, column provided for information purposes
Ib	Number of events in intervention group, before treatment	Integer	–	Not used in any of the methods, column provided for information purposes and/or future use
Ia	Number of events in intervention group, after treatment	Integer	Yes* (1a, 1b)	
Cb	Number of events in control group, before treatment	Integer	–	Not used in any of the methods, column provided for information purposes and/or future use
Ca	Number of events in control group, after treatment	Integer	Yes* (1a, 1b)	
$mean(Ib)$	Mean effect, intervention group, before treatment	Real	–	Not used in any of the methods, column provided for information purposes and/or future use
$mean(Ia)$	Mean effect, intervention group, after treatment	Real	Yes* (3, 4, 5, 6)	
$mean(Cb)$	Mean effect, control group, before treatment	Real	–	Not used in any of the methods, column provided for information purposes and/or future use
$mean(Ca)$	Mean effect, control group, after treatment	Real	Yes* (3, 4, 5, 6)	
$SD(Ia)$	Standard deviation of the effect, intervention group, after treatment	Real	Yes* (4)	
$SD(Ca)$	Standard deviation of the effect, control group, after treatment	Real	Yes* (4)	
MD	Means difference $MD = mean(Ia) - mean(Ca)$	Real	Yes* (3, 4, 5, 6)	If MD has not been inputted it is calculated by the formula. If it has, the inputted value is used instead (to take into account adjusted MD values)
$lCI95(MD)$	Lower limit of 95% Confidence Interval for the means difference.	Real	Yes* (3)	
$uCI95(MD)$	Upper limit of 95% Confidence Interval for the means difference.	Real	Yes* (3)	
$median(Ia)$	Median of effect, intervention group, after treatment	Real	–	Not used in any of the methods, column provided for information purposes and/or future use

table continued to next page

Name	Label	Type	Required	Information
$median(Ca)$	Median of effect, control group, after treatment	Real	–	Not used in any of the methods, column provided for information purposes and/or future use
$lCI95(Ia)$	Lower limit of 95% Confidence Interval for the mean of the intervention group, after treatment	Real	Yes* (5)	
$uCI95(Ia)$	Upper limit of 95% Confidence Interval for the mean of the intervention group, after treatment	Real	Yes* (5)	
$lCI95(Ca)$	Lower limit of 95% Confidence Interval for the mean of the intervention group, after treatment	Real	Yes* (5)	
$uCI95(Ca)$	Upper limit of 95% Confidence Interval for the mean of the intervention group, after treatment	Real	Yes* (5)	
OR	Odds ratio: $OR = \frac{\frac{Ia}{NIa - Ia}}{\frac{Ca}{NCa - Ca}}$	Real	Yes* (2)	
$lCI95(OR)$	Lower limit of 95% Confidence Interval for after treatment Odds Ratio	Real	Yes* (2)	
$uCI95(OR)$	Upper limit of 95% Confidence Interval for after treatment Odds Ratio	Real	Yes* (2)	
p-value	P-value of a two way test that compares between groups	Real	Yes* (6, 7)	
t-value	T-value of a two way t-test that compares between groups	Real	Yes* (6, 7)	A p-value is calculated using this value, which always overrides an inputted p-value in the previous field
df	Degrees of freedom of a two way t-test that compares between groups	Integer	Yes* (6, 7)	Degrees of freedom are automatically computed (if not provided) as $NIa + NCa + 2$

table continued to next page

Name	Label	Type	Required	Information
SEdiff	Standard Error of Difference between the means of the two groups	Real	Yes* (3)	
direction	Direction of the effect	Char	Yes	Leave empty if effect favours intervention. Input a single minus sign to reverse effect, if it favours control
quality	Quality of the study	Integer	No	Evaluation of each study (only needs to be inputted once for each study) For future use and information purposes only.
subgroup	Subgroup information for an outcome	String	No	Information on subgroup outcomes. They are used to label outcomes in the results.

Table 1: Fields in the data-entry worksheet. Fields marked with an asterisk are required for an effect size calculation method or more, but not all. The method(s) involved are shown in brackets.

3. Effect size and standard error calculation methods

A description of the eight methods that are used for the effect size and *SE* calculations is provided in this section, all inferred from The Cochrane Collaboration Handbook for Systematic Reviews of Interventions version 4.2.6, Section 8.5 (Higgins and Green 2006). Methods have been labelled 1a, 1b, 2, 3, 4, 5, 6, and 7. The first three deal with dichotomous data, the next four with continuous and the last method applies to both types. Each method can compute an effect size and variance from a different combination of available statistical parameters (sample sizes, means, t-values, etc). Table 2 provides information on the specific parameters that are needed as input for the application of each method.

The items that appear on Table 2 are described below:

- NIa is the size of the intervention group.
- NCa is the size of the control group.
- Ia is the number of events in the intervention group (always less than NIa).
- Ca is the number of events in the control group (always less than NCa).
- OR is the Odds Ratio: $OR = \dfrac{Ia/(NIa - Ia)}{Ca/(NCa - Ca)}$.
- $CI95(OR)$ is the 95% Confidence Interval for the Odds Ratio (fields $lCI95(OR)$ and $uCI95(OR)$ in the data sheet).
- MD is the means difference of the two groups, either provided or calculated with: $MD = mean(Ia) - mean(Ca)$.
- $CI95(MD)$ is the 95% Confidence Interval for the means difference (fields $lCI95(MD)$ and $uCI95(MD)$ in the data sheet).

Method	Data type	Input parameters needed	Effect estimate measure	Priority
1a	Dichotomous	*NIa* and *NCa* and *Ia* and *Ca*	Risk Difference	D2
1b	Dichotomous	*NIa* and *NCa* and *Ia* and *Ca*	Odds Ratio	D3
2	Dichotomous	*NIa* and *NCa* and *OR* and *CI95*(*OR*)	Odds Ratio	D1
3	Continuous	*NIa* and *NCa* and *MD* and (*CI95*(*MD*) or *SEdiff*)	Mean Difference	C3
4	Continuous	*NIa* and *NCa* and *MD* and *SD*(*Ia*) and *SD*(*Ca*)	Mean Difference	C1
5	Continuous	*NIa* and *NCa* and *MD* and *CI95*(*Ia*) and *CI95*(*Ca*)	Mean Difference	C2
6	Continuous	*NIa* and *NCa* and *MD* and (p-value or (t-value and *df*))	Mean Difference	C4
7	Continuous or Dichotomous	*NIa* and *NCa* and (p-value or (t-value and *df*))	Any	D4/C5

Table 2: Effect calculation methods by input parameters needed. Method priority is described in the text.

- *SEdiff* is the Standard Error of Difference between the means of the two groups.
- *SD*(*Ia*) is the Standard Deviation for the intervention group.
- *SD*(*Ca*) is the Standard Deviation for the control group.
- *CI95*(*Ia*) is the 95% Confidence Interval for the mean of the intervention group (fields *lCI95*(*Ia*) and *uCI95*(*Ia*) in the data sheet).
- *CI95*(*Ca*) is the 95% Confidence Interval for the mean of the control group (fields *lCI95*(*Ca*) and *uCI95*(*Ca*) in the data sheet).
- p-value is the p-value of the test.
- t-value is the t-value of the test.
- *df* the degrees of freedom of the t-test.

For some outcomes enough data is provided for the application of more than one of the methods. In such cases, the effect size and *SE* are calculated using all possible 'options' which enables the user to compare the results and the accuracy of the information supplied by the study in question (Figure 4). Nevertheless, only one method is finally selected to provide us with effect sizes and standard errors for the plots and the meta-analysis. The methods have been prioritised according to expected precision: that is, the expectation that the effect size and associated variance computed from the input data will be accurate. As a general rule, the fewer the number of mathematical transformations involved in getting from the 'raw data' to the statistical parameters used as input for the method, the higher the expected precision. Priority orders are displayed in Table 2 (e.g., D1 refers to the first

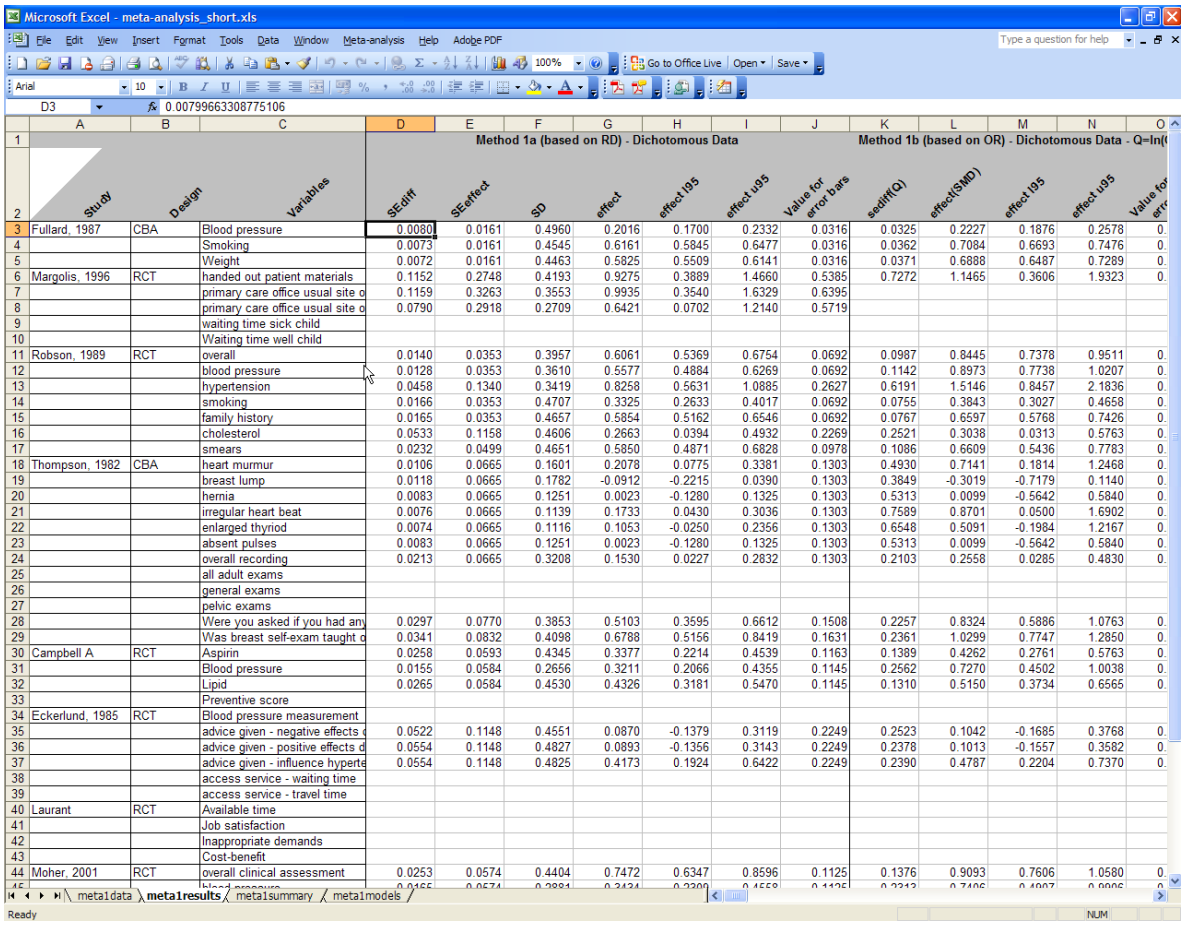


Figure 4: Results sheet.

choice when data is dichotomous, C3 to the third option when data is continuous etc) and the 'best' available option is automatically selected. In many studies, as a result of rounding etc the reported p-value is not very precise; therefore methods 6 and 7 are only used when no other method can be employed. Since 1a and 1b require the exactly same input, the former is arbitrarily prioritised over the latter, which - in effect - is never used in further analyses but is provided for comparison. The methods are described in detail in Appendix A.

4. Results summary

Once effect sizes and standard errors have been computed, we select the most precise result available for each outcome and summarise the results using a forest plot type graph (Figure 5). Forest plots normally display information on a single outcome from each study along with an overall effect estimate at the bottom of the plot. However, in our summary forest plot all outcomes for which an effect and a *SE* could be computed are included, so that the reader can perceive a general overview of all the outcomes of interest (Figure 6).

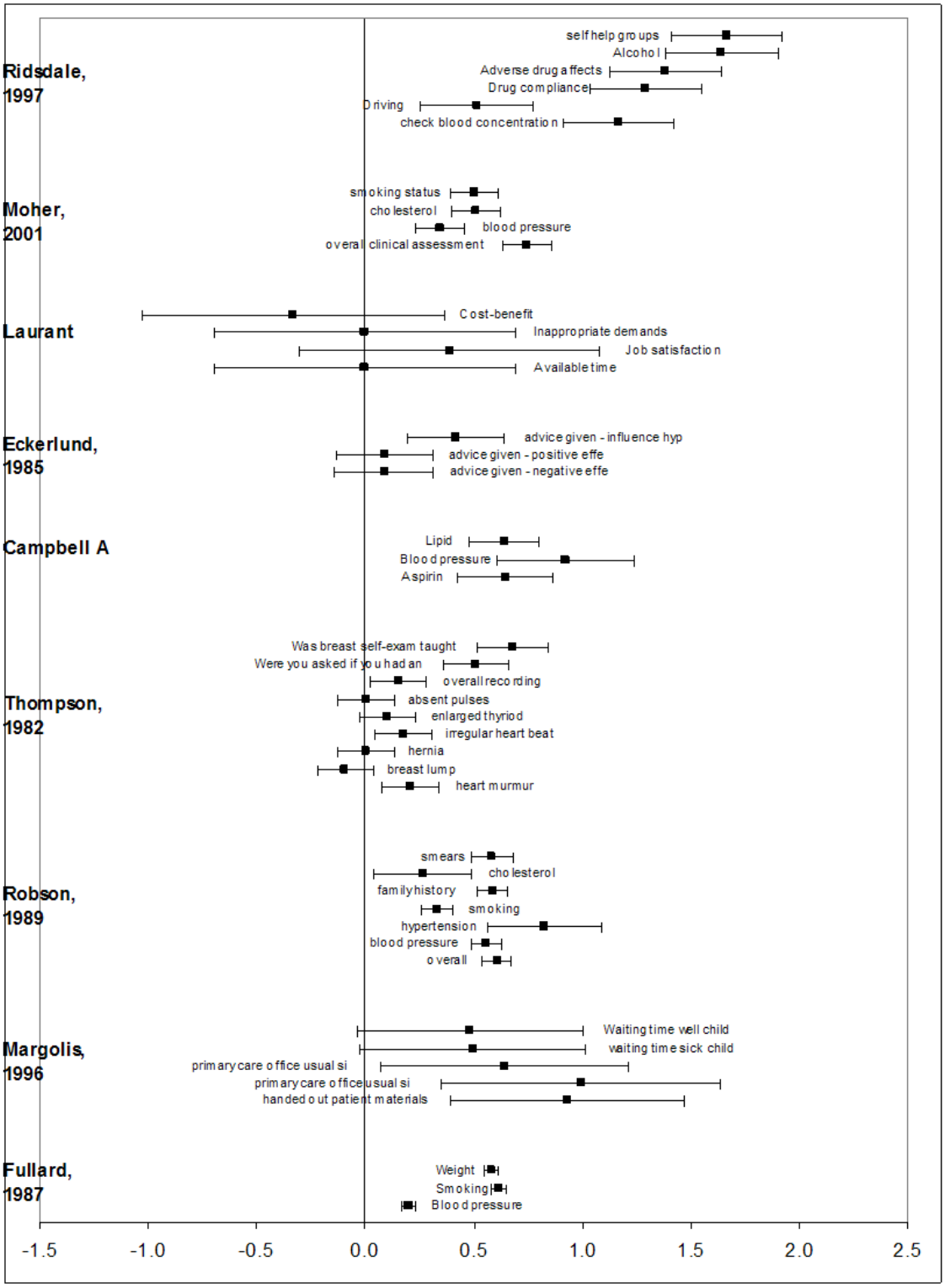


Figure 5: Forest plot on all outcomes.

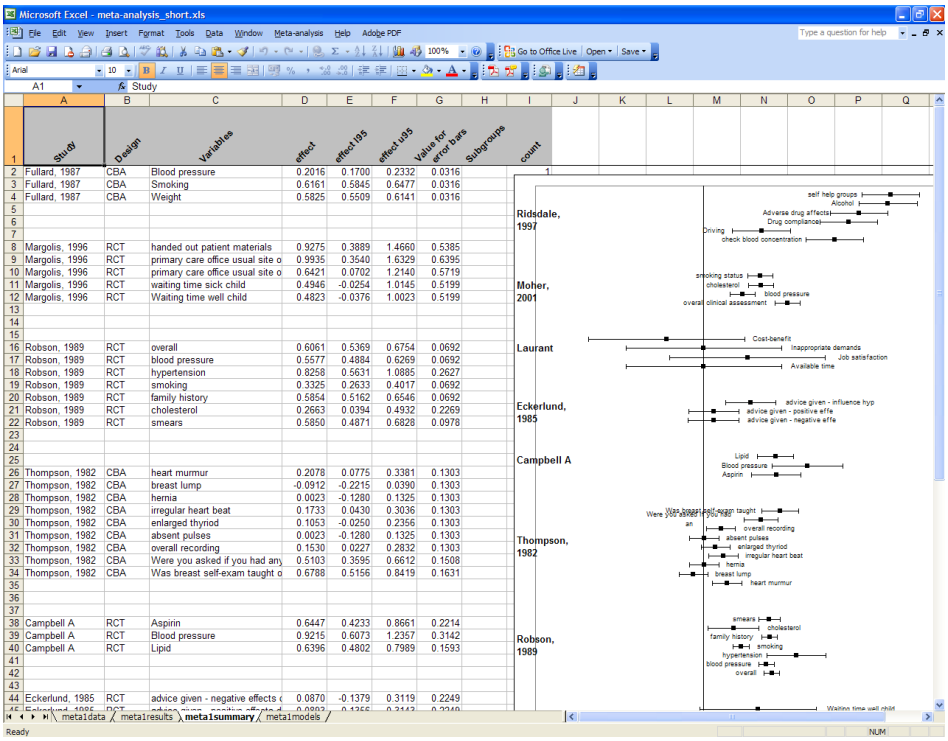


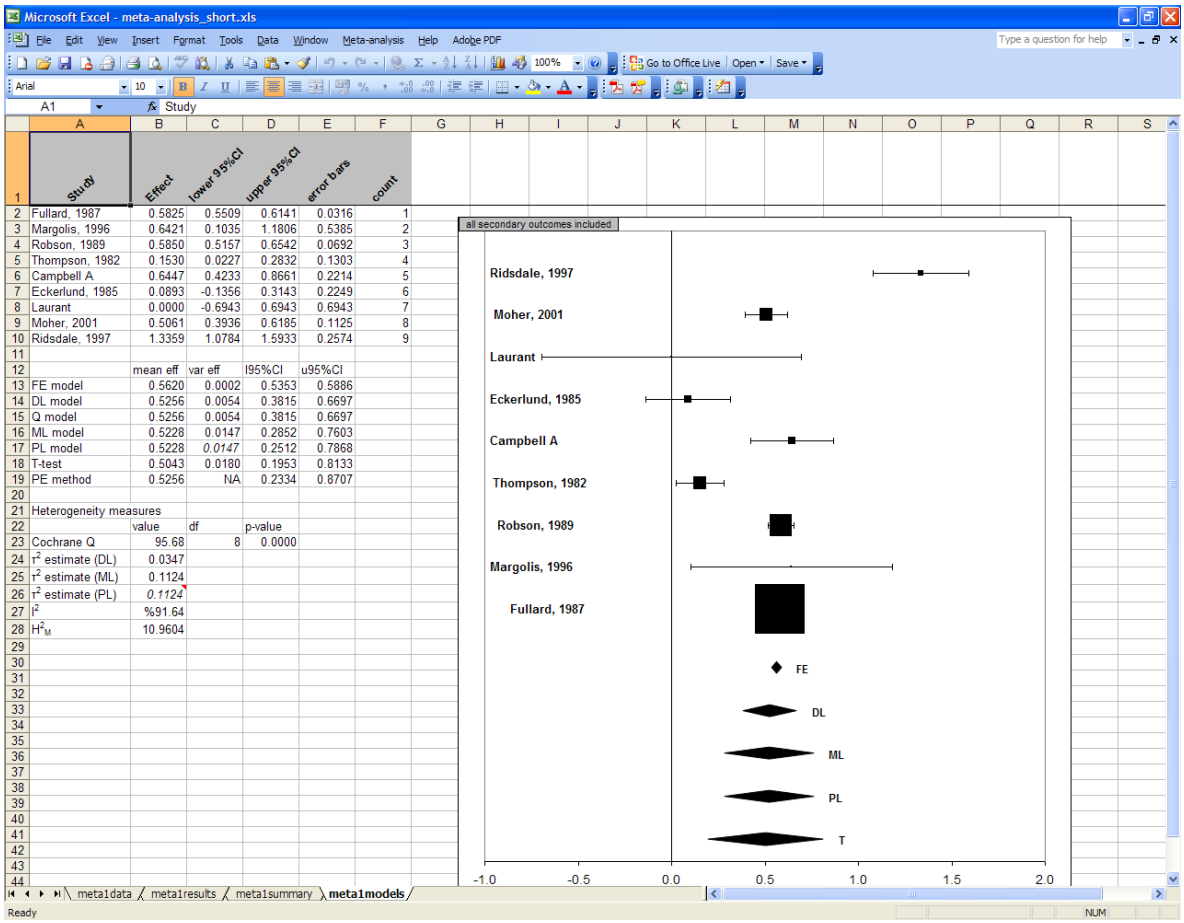
Figure 6: Summary sheet.

5. Meta-analysis methods overview

A primary concern for meta-analysts is statistical heterogeneity in true effect size across the studies included in an analysis, often attributed to clinical and/or methodological diversity (Higgins and Green 2006). More specifically, clinical heterogeneity describes variability that arises from different populations, interventions, outcomes and follow-up times, while methodological heterogeneity is related to differences in trial design and quality (Thompson 1994).

If the variation among the evaluated true effects of individual studies is not above that expected by chance (homogeneity not rejected) researchers usually select the fixed-effect model (Brockwell and Gordon 2001) to combine the separate estimates into a single result. However, medical research studies, even into the same issue, can vary on so many factors that homogeneity is a rare commodity and some degree of variability between studies may be anticipated (Thompson and Pocock 1991). The best viable approach in such cases is to summarize the results of the heterogeneous studies using a random-effects model (DerSimonian and Laird 1986). Models of this family take into account the identified between-study variation, estimate it and generally produce wider confidence intervals for the overall effect than a fixed-effects analysis.

As an alternative to frequentist random-effects meta-analysis, a researcher may choose a Bayesian approach to estimation of the between-study variance (Smith, Spiegelhalter, and Thomas 1995; Sutton and Abrams 2001). Unlike most conventional random effects methods Bayesian methods attempt to model the uncertainty in the between-study variance, but require specification of a prior distribution, a sometimes difficult task that is particularly crucial



	Study	Effect	lower 95%CI	upper 95%CI	error bars	count
2	Fullard, 1987	0.5825	0.5509	0.6141	0.0316	1
3	Margolis, 1996	0.6421	0.1035	1.1806	0.5385	2
4	Robson, 1989	0.5850	0.5157	0.6542	0.0692	3
5	Thompson, 1982	0.1530	0.0227	0.2832	0.1303	4
6	Campbell A	0.6447	0.4233	0.8661	0.2214	5
7	Eckerlund, 1985	0.0893	-0.1356	0.3143	0.2249	6
8	Laurant	0.0000	-0.6943	0.6943	0.6943	7
9	Moher, 2001	0.5061	0.3936	0.6185	0.1125	8
10	Ridsdale, 1997	1.3359	1.0784	1.5933	0.2574	9
12		mean eff	var eff	l95%CI	u95%CI	
13	FE model	0.5620	0.0002	0.5353	0.5886	
14	DL model	0.5256	0.0054	0.3815	0.6697	
15	Q model	0.5256	0.0054	0.3815	0.6697	
16	ML model	0.5228	0.0147	0.2852	0.7603	
17	PL model	0.5228	*0.0147*	0.2512	0.7868	
18	T-test	0.5043	0.0180	0.1953	0.8133	
19	PE method	0.5256	NA	0.2334	0.8707	
21	Heterogeneity measures					
22		value	df	p-value		
23	Cochrane Q	95.68	8	0.0000		
24	τ^2 estimate (DL)	0.0347				
25	τ^2 estimate (ML)	0.1124				
26	τ^2 estimate (PL)	*0.1124*				
27	I^2	%91.64				
28	H^2_M	10.9604				

Figure 7: Models sheet.

when the number of studies included in the meta-analysis is small Lambert, Sutton, Burton, Abrams, and Jones (2005).

We have implemented seven frequentist meta-analysis methods in the add-in, for calculating a mean effect estimate and its confidence intervals (Figure 7): t-test (T), fixed-effects model (FE), DerSimonian and Laird random-effects model (DL), Q-based method (Q), maximum-likelihood random-effects model (ML), profile-likelihood random-effects model (PL) and permutations method utilising a DL random-effects model (PE). Where multiple primary outcomes are available for a study, their median is used in the models. If no primary outcomes are available, the secondary outcomes median is used instead.

In addition, a variety of heterogeneity measures are provided to help the user decide on the fixed or random effects approach (either DL, PL or ML; PE and T are special cases). Cochran's Q provides a p-value for the test of homogeneity, when compared with a χ^2_{k-1} distribution (Brockwell and Gordon 2001), where k is the number of studies. However the test is known to be poor at detecting heterogeneity since its power is low when the number of studies is small (Hardy and Thompson 1998). I^2 is deemed to be more reliable in assessing inconsistency between studies, with values of 25%, 50% and 75% corresponding to low, moderate and high heterogeneity respectively (Higgins, Thompson, Deeks, and Altman 2003). H^2_M is the

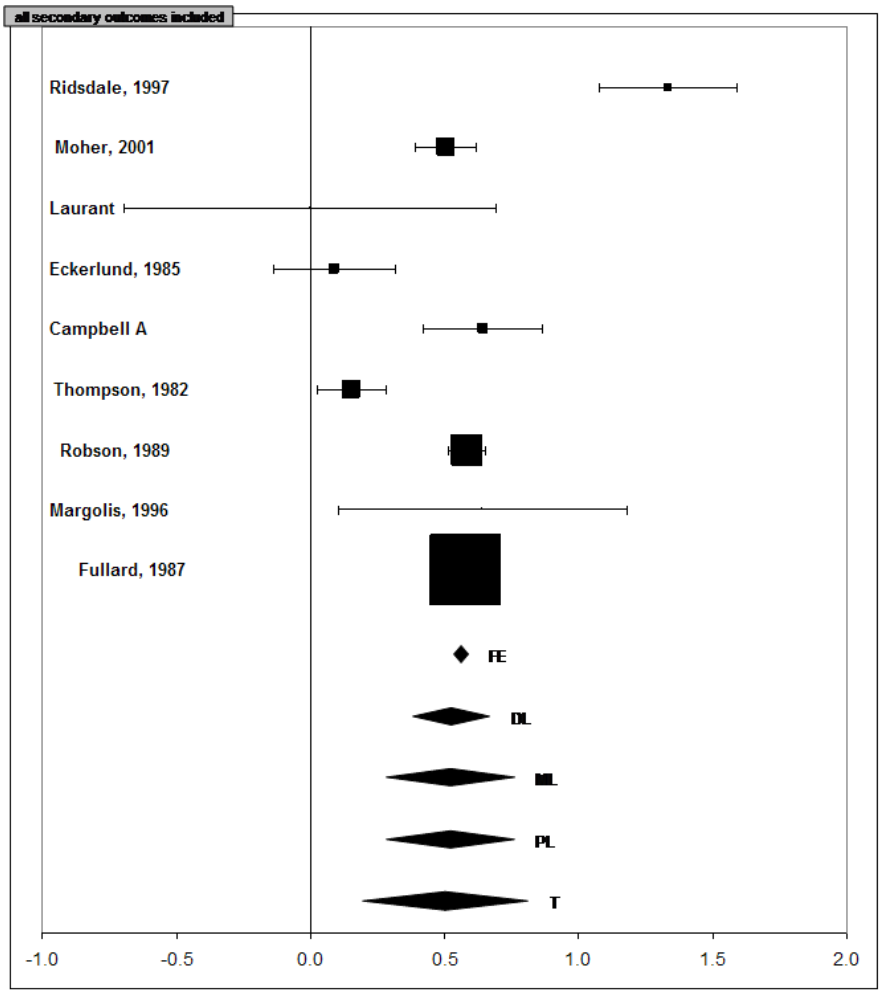


Figure 8: Forest plot on median outcomes and meta-analysis models.

only measure that is truly independent of k and it takes values in the $[0, +\infty)$ range with 0 indicating - a somewhat worrying - perfect homogeneity (Mittlbock and Heinzl 2006). It has also been suggested that the estimate of between-study variance, $\hat{\tau}^2$, should be reported (Higgins and Green 2008), therefore the estimated between-study variance, where relevant to the model, is also displayed.

The MA methods and heterogeneity measures along with certain prerequisite assumptions are described in the following section, while the resulting forest plot (which again uses the median for studies with multiple primary outcomes) can be observed in Figure 8.

6. Meta-analysis methods description

6.1. General

Let us consider a group of studies, whose overall effect we wish to estimate. The simplest

way to estimate a confidence interval for the population mean is to treat the effect for each study as a single observation from a population of studies, and apply the usual Normal or T distribution assumptions.

$$\bar{X} \pm z\frac{s}{\sqrt{k}} \tag{1}$$

$$\bar{X} \pm t_{n-1}\frac{s}{\sqrt{k}} \tag{2}$$

where $\bar{X}$ is the sample mean of the effect sizes, s their standard deviation and z, t_{n-1} are the critical values for the two distributions.

This method, although simple, is thought to be deficient compared to the more advanced techniques since it does not take into account the variances on the effect estimates within the studies - provided they have been reported, which is not always the case. Nevertheless, it has been evaluated as a reliable alternative in performing meta-analyses, particularly if control for type I error rate is important (Follmann and Proschan 1999). This finding was supported by a simulation study that compared the performance of meta-analysis methods across effects of various distributions (Kontopantelis and Reeves 2008).

6.2. Fixed effects

If the within study variances are available, a researcher can take these into account in the estimation of an overall effect μ and a confidence interval. The fixed-effects model assumes that within-study variances may differ, but that there is homogeneity in effect size across all studies. It can be defined as:

$$Y_i = \theta_i + e_i, \quad e_i \sim N(0, \sigma_i^2) \tag{3}$$

where, for study i: Y_i is the effect size estimate, θ_i the true effect size and e_i the random error. Under the fixed effects approach, the true study effects are all assumed to be equal ($\theta_i = \mu, \quad i = 1, 2, \cdots, k$) and the only deviations from the true effect size occur because of errors e_i, representing the imprecision of results within each study. Those errors are assumed to be independent and normally distributed with mean zero and variance σ_i^2, therefore the estimated effect sizes Y_i are also normally distributed with variance σ_i^2, but with mean θ_i. The fixed effect estimate of the overall effect μ is usually calculated as a weighted average, using the within study variances σ_i^2 as precision weights (Brockwell and Gordon 2001):

$$\hat{\mu}_F = \frac{\sum_{i=1}^{k} \hat{w}_i Y_i}{\sum_{i=1}^{k} \hat{w}_i} \tag{4}$$

where $\hat{w}_i = 1/\hat{\sigma}_i^2$ and $var(\hat{\mu}_F) = \dfrac{1}{\sum_{i=1}^{k} \hat{w}_i}$.

6.3. Random effects

Often the homogeneity assumption is unlikely and variation in the true effect across studies is assumed, which is attributed to differences in design, circumstances, populations of participants and dose/type of treatment offered (Van den Noortgate and Onghena 2003). Whenever heterogeneity is identified with Cochran's Q - test (Cochran 1937) or other appropriate statistical measures like I^2 or H_M^2, a random effects model may be preferred over the fixed effects

model. More specifically $I^2 = 100\dfrac{Q - df}{Q}$ (Higgins and Thompson 2002) and $H_M^2 = \dfrac{Q - df}{df}$ (Mittlbock and Heinzl 2006) are both measures that use the relation of the between- to the within-study variance to evaluate heterogeneity.

The random-effects approach assumes that true effects θ_i are normally distributed. This incorporates a second error term in (3), which accounts for across study variability, and the model becomes (Brockwell and Gordon 2001):

$$Y_i = \theta_i + e_i, \quad e_i \sim N(0, \sigma_i^2) \tag{5}$$

$$\theta_i = \mu + \epsilon_i, \quad \epsilon_i \sim N(0, \tau^2) \tag{6}$$

In this case, the overall effect estimate is provided by:

$$\hat{\mu}_R = \frac{\sum_{i=1}^{k} \hat{w}_i' Y_i}{\sum_{i=1}^{k} \hat{w}_i'} \tag{7}$$

where $\hat{w}_i' = \dfrac{1}{(\tau^2 + \hat{\sigma}_i^2)}$ and $var(\hat{\mu}_R) = \dfrac{1}{\sum_{i=1}^{k} \hat{w}_i'}$. This model, by assuming that the studies at hand are a random sample out of a population of studies and modelling the variability between them, provides wider confidence intervals for the estimate of the overall effect and is generally considered to be a more conservative approach than the fixed effects model (Poole and Greenland 1999).

The variance parameter τ^2, of the between studies error term ϵ_i, is a measure of the between study heterogeneity. Since it is rarely - if ever - known, it needs to be estimated. DerSimonian and Laird's estimator for τ^2 is the method most widely used (DerSimonian and Laird 1986). However, the overall effect and the between study variance can also be estimated with methods such as the simple maximum likelihood and the profile likelihood proposed by Hardy and Thompson (Hardy and Thompson 1996). The simple maximum likelihood estimates $\hat{\mu_{mi}}$ and $\hat{\tau_{mi}^2}$ can be obtained from the log-likelihood function in (8).

$$\log L(\mu, \tau^2) = -\frac{1}{2}\sum_{i=1}^{k} \log(2\pi(\hat{\sigma}_i^2 + \tau^2)) - -\frac{1}{2}\sum_{i=1}^{k} \frac{(y_i - \mu)^2}{\hat{\sigma}_i^2 + \tau^2}, \quad \mu \in \Re \ \& \ \tau^2 \geq 0 \tag{8}$$

The log-likelihood function is solved iteratively for the simple maximum likelihood method, and the estimated variances are generally more accurate than the ones provided by the DL method. The profile likelihood method is also iterative, but provides confidence intervals for the estimates by taking into account the fact that when estimating one parameter the other one is also unknown and needs to be estimated as well. Computationally, it can be described as more advanced version of maximum likelihood which converges to a solution using nested iterations. Simulations show that this produces more accurate confidence intervals than other methods (Kontopantelis and Reeves 2008). For more details on these methods see Brockwell and Gordon (2001).

Another proposed random-effects method has been described by Follmann and Proschan (1999), which we have conveniently labelled the permutations method. The method can be described in three steps. First, a matrix of all possible combinations of study outcomes from a meta-analysis is created by permuting the effects signs. Then the DL method is used to compute the overall effect for each combination. Finally, using the created distribution for

the overall effect estimate $\hat{\mu_{pe}}$ and the hypothesis that μ is zero, a Confidence Interval for the overall effect is calculated. In other words, rejection (or not) of the hypothesis is decided on the position of the observed mean estimate within the distribution.

Since more often than not Cochran's Q-test is the main determinant in the fixed or random-effects model selection, we also include a mixed approach that combines these two models according to the outcome of the Q-test. If the homogeneity hypothesis is rejected, the DL random-effects model is selected; otherwise we choose the fixed-effects model. In accordance with Brockwell and Gordon (2001) we label this method 'Q'. All models (except T and PE, for which it does not apply) use the inverse variance method to calculate the overall effect and standard error of a meta-analysis.

7. Validation

Methods and graphs were validated using eight published meta-analyses whose full datasets were readily available (Laurant, Reeves, Kontopantelis, Hermens, Braspenning, Grol, Wensing, and Sibbald 2008). Validation was performed with Stata version 9.2 (StataCorp. 2007) and more specifically programs `metan` version 2.34 (Harris, Bradburn, Deeks, Harbord, Altman, and Sterne 2008) and `meta` version 2.06 (Sharp and Stern 1997). Results agreed completely for the methods that are available in both platforms: fixed-effect model (inverse variance), random-effects model (DerSimonian and Laird, inverse variance), Cochran's Q and I^2. Comparisons for the more advanced maximum likelihood, profile likelihood and permutation methods could not be performed, since they are not included in any of the statistical packages considered in this paper.

8. Using MetaEasy

The add-in is freely available for download from the National Primary Care Research and Development Centre's web page (`http://www.npcrdc.ac.uk/`) and the first author's personal web page (`http://www.statanalysis.co.uk/`). It is provided as an installer executable file, along with a short manual and data examples. Once it has been installed a meta-analysis menu option will always be available on the menu bar (Figure 3). The first step in creating a meta-analysis workbook is to create and format the worksheets appropriately. This is accomplished with the Format Sheets command which creates five worksheets for a single meta-analysis. All data entry is done on the first sheet: the remaining four sheets are 'locked' and are where the results and graphs are output.

After the data has been input, obtaining the results is an one click process, using one of the two available options: (i) include all primary and secondary outcomes in the meta-analysis models or (ii) only include secondary outcomes where a study has no primary outcomes. Studies frequently collect, analyse, and report data on a number of outcomes relevant to a given meta-analysis. Usually a well-designed study provides a precise description and definition of objectives and outcomes. This allows the outcomes to be classified as either primary or secondary: the former address the focus of the study while the latter allow for the investigation of subsidiary questions (Gebski, Marschner, and Keech 2002). If the study is not clear about which outcomes were regarded primary, then they should all be treated as secondary.

In addition, the reported heterogeneity measures help the user select between the fixed-effects

and a random-effects model. Details on the characteristics and usage of the measures have been provided in Sections 5 and 6.

Since copying and pasting the results from **Excel** to a text editor is not always fully successful, graphs and blocks of cells can be exported as Graphics Interchange Format (`GIF`) files using the Export graph command. The code for this command, which can be used independently to export any graphs and cell ranges, was obtained from various websites and authors (Staff 2004; McRitchie, Bullen, and Peltier 2008) then edited and integrated into the add-in. However, the code was created in the Microsoft **Office** 2003 suite and it has not yet been tested with other versions.

9. Conclusions

In this article we have described an efficient way to perform meta-analyses in **Excel**. Some of the advantages of our module over established statistical and meta-analysis software packages include:

- Ease of use.
- Time saving, since effect sizes and standard errors are calculated automatically from whatever statistical parameters are available, without the need for prior transformation to a common metric.
- More robust calculation of the effect sizes and *SE*, since the 'best' method is automatically selected.
- The extracted data from each study are easily accessible, they can be quickly edited or corrected and the analysis repeated.
- Includes a choice of seven meta-analysis models (instead of the usual two).
- Provides three advanced methods, maximum likelihood, profile likelihood and permutation, not currently available in other software packages.
- Effect sizes and standard errors can be exported for use in other MA software packages.
- Provides a descriptive forest plot allowing multiple outcomes per study.
- Provides the final forest plot with the estimated mean effect and confidence interval.
- Reports a wide range of heterogeneity measures.
- The graphs can be easily edited and exported.
- It is free (provided Microsoft **Excel** is available).

We will aim to provide additional tools with future versions of the module (e.g., a funnel plot, Mantel-Haenszel method for binary data), as suggested by user feedback.

Acknowledgments

We would like to thank the two anonymous referres, for their helpful and observant comments.

References

Bailar JC (1997). "The Promise and Problems of Meta-Analysis." *New England Journal of Medicine*, **337**(8), 559–561.

Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KGM (2006). "Development and Validation of **MIX**: Comprehensive Free Software for Meta-Analysis of Causal Research Data." *BMC Medical Research Methodology*, **6**(50).

Brockwell SE, Gordon IR (2001). "A Comparison of Statistical Methods for Meta-Analysis." *Statistics in Medicine*, **20**(6), 825–840.

Cochran WG (1937). "Problems Arising in the Analysis of a Series of Similar Experiments." *Journal of the Royal Statistical Society*, **4**(1), 102–118. Supplement.

DerSimonian R, Laird N (1986). "Meta-Analysis in Clinical Trials." *Controled Clinical Trials*, **7**(3), 177–188.

Egger M, Smith GD (1997). "Meta-Analysis. Potentials and Promise." *British Medical Journal*, **315**(7119), 1371–1374.

Egger M, Smith GD, Phillips AN (1997). "Meta-Analysis: Principles and Procedures." *British Medical Journal*, **315**(7121), 1533–1537.

Egger M, Sterne J, Smith G (1998). "Meta-Analysis Software." *British Medical Journal*, **316**(7126). Complementary article, URL http://www.bmj.com/archive/7126/7126ed9.htm.

Eysenck HJ (1994). "Meta-Analysis and its Problems." *British Medical Journal*, **309**(6957), 789–792.

Follmann DA, Proschan MA (1999). "Valid Inference in Random Effects Meta-Analysis." *Biometrics*, **55**(3), 732–737.

Gebski V, Marschner I, Keech AC (2002). "Specifying Objectives and Outcomes for Clinical Trials." *Medical Journal of Australia*, **176**(10), 491–492.

Glass GV (1976). "Primary, Secondary, and Meta-Analysis of Research." *Educational Researcher*, **5**(10), 3–8.

Hardy RJ, Thompson SG (1996). "A Likelihood Approach to Meta-Analysis with Random Effects." *Statistics in Medicine*, **15**(6), 619–629.

Hardy RJ, Thompson SG (1998). "Detecting and Describing Heterogeneity in Meta-Analysis." *Statistics in Medicine*, **17**(8), 841–856.

Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J (2008). "**Metan**: Fixed- and Random-Effects Meta-Analysis." *Stata Journal*, **8**(1), 3–28.

Higgins JP, Green S (2006). *Cochrane Handbook for Systematic Reviews of Interventions: Version 4.2.6*. URL http://www.cochrane.org/resources/handbook/Handbook4.2.6Sep2006.pdf.

Higgins JP, Green S (2008). *Cochrane Handbook for Systematic Reviews of Interventions: Version 5.0.1*. URL http://www.cochrane-handbook.org/.

Higgins JP, Thompson SG (2002). "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine*, **21**(11), 1539–1558.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003). "Measuring Inconsistency in Meta-Analyses." *British Medical Journal*, **327**(7414), 557–560.

Huque MF (1988). "Experiences with Meta-Analysis in NDA Submissions." *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, **2**, 28–33.

Kontopantelis E, Reeves D (2008). "A Comparison of Random Effects Meta-Analysis Methods when Study Effects are Non-Normally Distributed." In *2008 Annual Conference of the Royal Statistics Society, Abstracts Booklet*, p. 52. RSS,12 Errol Street, London, EC1Y 8LX. URL http://www.rss.org.uk/pdf/RSS2008_Final_abstracts_booklet.pdf.

Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR (2005). "How Vague is Vague? A Simulation Study of the Impact of the Use of Vague Prior Distributions in MCMC Using **WinBUGS**." *Statistics in Medicine*, **24**(15), 2401–2428.

Laurant M, Reeves D, Kontopantelis E, Hermens R, Braspenning J, Grol R, Wensing M, Sibbald B (2008). "The Effectiveness of Nurse Supplementation in Primary Care: A Systematic Review and Meta-Analysis." Unpublished.

Lewis S (2001). "Forest Plots: Trying to See the Wood and the Trees." *British Medical Journal*, **322**(7300), 1479–1480.

McRitchie D, Bullen S, Peltier J (2008). "The **MakeGIF** Utility." URL http://www.ac6la.com/makegif.html.

Mittlbock M, Heinzl H (2006). "A Simulation Study Comparing Properties of Heterogeneity Measures in Meta-Analyses." *Statistics in Medicine*, **25**(24), 4321–4333.

Poole C, Greenland S (1999). "Random-Effects Meta-Analyses Are Not Always Conservative." *American Journal of Epidemiology*, **150**(5), 469–475.

Rosenberg MS, Adams DC, Gurevitch J (2007). ***MetaWin*** *Statistical Software*. Sunderland, MA. URL http://www.metawinsoft.com/.

SAS Institute Inc (2003). *The SAS System, Version 9.1*. Cary, NC. URL http://www.sas.com/.

Sharp S, Stern J (1997). "Meta-Analysis." *Stata Technical Bulletin Reprints*, **7**, 100–106.

Smith TC, Spiegelhalter DJ, Thomas A (1995). "Bayesian Approaches to Random-Effects Meta-Analysis: A Comparative Study." *Statistics in Medicine*, **14**(24), 2685–2699.

SPSS Inc (2006). *SPSS for Windows, Release 14*. SPSS Inc., Chicago, IL. URL http://www.spss.com/.

Staff H (2004). "**XL2GIF** Routine and Manipulations." URL http://www.mvps.org/dmcritchie/excel/xl2gif.htm.

StataCorp (2007). *Stata Statistical Software: Release 10*. StataCorp LP, College Station, TX. URL http://www.stata.com/.

Sutton AJ, Abrams KR (2001). "Bayesian Methods in Meta-Analysis and Evidence Synthesis." *Statistical Methods in Medical Research*, **10**(4), 277–303.

The Cochrane Collaboration (2008). *Review Manager (**RevMan**) for Windows: Version 5.0*. The Nordic Cochrane Centre, Copenhagen. URL http://www.cc-ims.net/revman/.

Thompson SG (1994). "Why Sources of Heterogeneity in Meta-Analysis Should be Investigated." *British Medical Journal*, **309**(6965), 1351–1355.

Thompson SG, Pocock SJ (1991). "Can Meta-Analyses be Trusted?" *The Lancet*, **338**(8775), 1127–1130.

Van den Noortgate W, Onghena P (2003). "Multilevel Meta-Analysis: A Comparison with Traditional Meta-Analytical Procedures." *Educational and Psychological Measurement*, **63**(5), 765–790.

A. Effect size and SE calculation methods

Needed	NIa, NCa, Ia, Ca
Step 1	$SE_{diff}(RD) = \sqrt{\frac{P_{Ia}(1-P_{Ia})}{NIa} + \frac{P_{Ca}(1-P_{Ca})}{NCa}}$ where $P_{Ia} = Ia/NIa$ & $P_{Ca} = Ca/NCa$ (or: $SE_{diff}(RD) = \frac{RD}{z}$ where $RD = \frac{Ia}{NIa} - \frac{Ca}{NCa}$ & $z = \lvert normsinv(P/2)\rvert$*)
Step 2	$SE_{effect} = \sqrt{\frac{1}{NIa} + \frac{1}{NCa}}$
Step 3	$SD = SE_{diff}/SE_{effect}$
Step 4	$effect = RD/SD$
Step 5	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 3: Method 1a (based on risk difference): Dichotomous data. * **Excel** function *normsinv* returns the inverse of the standard normal cumulative distribution and P refers to the p-value of the test. This alternative approach is never used since it is less precise (only provided for completeness).

Needed	NIa, NCa, Ia, Ca
Step 1	Let $Q = \ln(OR)$, where $OR = \frac{Ia/(NIa-Ia)}{Ca/(NCa-Ca)}$*
Step 2	$SE_{diff}(Q) = \sqrt{\frac{1}{Ia} + \frac{1}{NIa-Ia} + \frac{1}{Ca} + \frac{1}{NCa-Ca}}$
Step 3	$upperCI95\%(Q) = Q + 1.96 \cdot SE_{diff}(Q)$ $lowerCI95\%(Q) = Q - 1.96 \cdot SE_{diff}(Q)$
Step 4	$effect = Q \cdot \sqrt{3}/\pi$
Step 5	$upperCI95\%(effect) = upperCI95\%(Q) \cdot \sqrt{3}/\pi$ $lowerCI95\%(effect) = lowerCI95\%(Q) \cdot \sqrt{3}/\pi$

Table 4: Method 1b (based on odds ratio): Dichotomous data. * OR cannot be computed when $NIa = Ia$ or $NCa = Ca$.

Needed	NIa, NCa, OR, $CI95\%(OR)$
Step 1	Calculate 'absolute' confidence intervals for Standardised Mean Difference: $upperCI95\%(SMD) = \frac{\sqrt{3}}{\pi} \ln upperCI95\%(OR)$ $lowerCI95\%(SMD) = \frac{\sqrt{3}}{\pi} \ln lowerCI95\%(OR)$
Step 2	$SE_{effect} = \frac{upperCI95\%(SMD) - lowerCI95\%(SMD)}{3.92}$ for $NIa \geq 60$ & $NCa \geq 60$ otherwise $SE_{effect} = \frac{upperCI95\%(SMD) - lowerCI95\%(SMD)}{2 \cdot tinv(1 - 0.95, NIa + NCa - 2)}$*
Step 3	$effect = \frac{\sqrt{3}}{\pi} \ln OR$ (effect is actually the SMD)
Step 4	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 5: Method 2 (based on odds ratio and its confidence interval): Dichotomous data. * **Excel** function *tinv* returns the t-value for specific alpha and degrees of freedom.

Needed	NIa, NCa, MD, $CI95\%(MD)$*
Step 1	$SE_{diff}(MD) = \frac{upperCI95\%(MD) - lowerCI95\%(MD)}{3.92}$ for $NIa \geq 60$ & $NCa \geq 60$ otherwise $SE_{diff}(MD) = \frac{upperCI95\%(MD) - lowerCI95\%(MD)}{2 \cdot tinv(1 - 0.95, NIa + NCa - 2)}$†
Step 2	$SE_{effect} = \sqrt{\frac{1}{NIa} + \frac{1}{NCa}}$
Step 3	$SD = SE_{diff} / SE_{effect}$
Step 4	$effect = MD/SD$
Step 5	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 6: Method 3 (based on mean difference and its confidence interval): Continuous data. * Instead of the $CI95\%(MD)$ the $SE_{diff}(MD)$ may be provided instead. † **Excel** function *tinv* returns the t-value for specific alpha and degrees of freedom.

Needed	NIa, NCa, MD, $SD(Ia)$, $SD(Ca)$*
Step 1	$SD = \sqrt{\frac{SD(Ia)^2 \cdot NIa + SD(Ca)^2 \cdot NCa}{NIa + NCa}}$
Step 2	$SE_{effect} = \sqrt{\frac{1}{NIa} + \frac{1}{NCa}}$
Step 3	$effect = MD/SD$
Step 4	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 7: Method 4 (based on mean difference and group variances): Continuous data. * Instead of $SD(Ia)$ & $SD(Ca)$ we may have $SEM(Ia)$ & $SEM(Ca)$ (Standard Error of Measurement). In that case we use $SEM = SD/\sqrt{N}$ to convert SEM values to SD ones.

Needed	NIa, NCa, MD, $CI95\%(Ia)$, $CI95\%(Ca)$
Step 1	$SD(Ia) = \sqrt{NIa}\frac{upperCI95\%(Ia) - lowerCI95\%(Ia)}{3.92}$ if $NIa \geq 60$ if $NIa < 60$ then $SD(Ia) = \sqrt{NIa}\frac{upperCI95\%(Ia) - lowerCI95\%(Ia)}{2 \cdot tinv(1 - 0.95, NIa - 1)}$*
Step 2	$SD(Ca) = \sqrt{NCa}\frac{upperCI95\%(Ca) - lowerCI95\%(Ca)}{3.92}$ if $NCa \geq 60$ if $NCa < 60$ then $SD(Ca) = \sqrt{NCa}\frac{upperCI95\%(Ca) - lowerCI95\%(Ca)}{2 \cdot tinv(1 - 0.95, NCa - 1)}$*
Step 3	$SD = SE_{diff}/SE_{effect}$
Step 4	$effect = MD/SD$
Step 5	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 8: Method 5 (based on mean difference and group confidence intervals): Continuous data. * **Excel** function *tinv* returns the t-value for specific alpha and degrees of freedom.

Needed	NIa, NCa, MD, p*
Step 1	$SE_{diff}(MD) = \dfrac{MD}{tinv(p, NIa + NCa - 2)}$†
Step 2	$SE_{effect} = \sqrt{\dfrac{1}{NIa} + \dfrac{1}{NCa}}$
Step 3	$SD = SE_{diff}/SE_{effect}$
Step 4	$effect = MD/SD$
Step 5	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 9: Method 6 (based on mean difference and p value): Continuous data. * A t-value and the degrees of freedom may be used instead of the p-value (in cases where comparison between groups was performed using a t-test). † **Excel** function *tinv* returns the t-value for specific alpha and degrees of freedom.

Needed	NIa, NCa, p*
Step 1	$z = \lvert normsinv(p/2)\rvert$†
Step 2	$SE_{effect} = \sqrt{\dfrac{1}{NIa} + \dfrac{1}{NCa}}$
Step 3	$effect = z \cdot SE_{effect}$
Step 4	$CI95\%(effect) = effect \pm 1.96 \cdot SE_{effect}$

Table 10: Method 7 (based on p value): Continuous or dichotomous data. * A t-value and the degrees of freedom may be used instead of the p-value (in cases where comparison between groups was performed using a t-test). † **Excel** function *normsinv* returns the inverse of the standard normal cumulative distribution.

Affiliation:

Evangelos Kontopantelis, David Reeves
National Primary Care Research and Development Centre (NPCRDC)
Faculty of Medical and Human Sciences
University of Manchester
Williamson Building 5th floor
Manchester M13 9PL, United Kingdom
E-mail: e.kontopantelis@manchester.ac.uk, david.reeves@manchester.ac.uk
URL: http://www.medicine.manchester.ac.uk/staff/EvanKontopantelis
http://www.medicine.manchester.ac.uk/staff/76561

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 30, Issue 7 *Submitted:* 2008-04-23
April 2009 *Accepted:* 2009-03-17